Biologically Plausible Reinforcement Learning

Rombouts, J.O.

2015

7 Summary

In this thesis we developed three models of reinforcement learning based on the framework of attention gated learning (Roelfsema and van Ooyen, 2005). In chapter 3, we reinterpreted the AGREL framework in terms of learning action-values. With this we showed how the framework can be applied to learning in settings with general rewards (compared to the $0-1$ rewards required in AGREL) and to settings where multiple simultaneous actions need to be selected (compared to the 1-of-c classification in AGREL) under the assumption of a scalar global prediction error signal. Preliminary work shows that the ideas in chapter 3 can be extended as to learn continuous actions with population coding output layers (Rombouts, Roelfsema, and Bohte, 2013). In chapter 4 we introduced AuGMEnT, which leaves the direct-reward setting of AGREL and MQ-AGREL for the delayed reward setting, thus showing how attention-gated learning can be used for learning action sequences, instead of single actions. The model is equipped with a working memory that can be used to learn difficult working memory tasks. The basic model can be seen as a biologically plausible implementation of the SARSA(λ) algorithm (Rummery and Niranjan, 1994), and the model with working memory extends this to learning problems in the class of partially observable MDPs (see chapter 2). The representations that the model learns are comparable to those that can be found in the brains of monkeys when they are trained on the same tasks. In chapter 5 we used AuGMEnT to model an attentional filtering task, investigating how such filtering can be learned by simple trial-and-error learning. One assumption we made for AuGMEnT is that a 'reset' signal is broadcast through the network after a trial ends, which can be seen as a supervised learning signal. In chapter 6 we extended AuGMEnT with an internal 'reset' action, and demonstrated that this extended model can learn trial boundaries, purely by reinforcement learning.